# IOOS Data Management Activities

Jeff de La Beaujardière

National Oceanic and Atmospheric Administration
Integrated Ocean Observing System (IOOS)
1100 Wayne Ave #1225
Silver Spring MD 20910 USA

*Abstract*- **The Integrated Ocean Observing System (IOOS) will provide information about open oceans and US coastal waters and Great Lakes to scientists, managers, businesses, governments, and the public in order to support research, to inform decision-making, and to enable new applications and derived products beyond the original intent of the data gathering. Programmatically, IOOS is a component of the Global Ocean Observing System (GOOS) and the Global Earth Observing System of Systems (GEOSS). The US National Oceanic and Atmospheric Administration (NOAA) has been assigned the role of lead federal agency in this endeavor. Technically, IOOS includes or interfaces with existing observing systems, data providers and archives, and collaborates in developing additional capabilities in observations, data management and data use. The NOAA IOOS office has established a limited-scope data-management prototype and is developing an architecture to extend and augment that prototype. This paper introduces a general architecture for IOOS and discusses specific standardization and implementation activities in that context.**

## I. INTRODUCTION

The Integrated Ocean Observing System (IOOS) will enhance our ability to collect, deliver, and use oceanographic information. The goal is to provide sustained data on our open oceans, coastal waters, and Great Lakes in the formats, rates, and scales required by scientists, managers, businesses, governments, and the public to support research and to inform decision-making. IOOS is the oceans-and-coasts component of the US Integrated Earth Observation System (IEOS), the US contribution to the Global Ocean Observing System (GOOS), and the US contribution to the oceans-and-coasts component of the Global Earth Observation System of Systems (GEOSS). In 2007, the US National Oceanic and Atmospheric Administration (NOAA) established an office (http://ioos.gov/) to manage its contributions to IOOS. In 2009, the US Congress passed the Integrated Coastal and Ocean Observation System Act which, among many other provisions, establishes NOAA as the lead federal agency for IOOS and calls for an Interagency Ocean Observation Committee to "establish protocols and standards for System data processing, management, and communication."[1]

In this paper we outline the conceptual architecture for IOOS and its data management component, we describe progress to date toward implementing parts of that architecture, and we discuss possible next steps.

## II. CONCEPTUAL IOOS ARCHITECTURE

The Integrated Ocean Observing System is in fact not a single system but a loose collection of systems, most of which the IOOS program does not control directly. The challenge is to overlay a distributed service-oriented architecture alongside existing functionality to simplify and expand data discovery, access, use and integration.

At the highest level, we divide IOOS into the five architectural layers shown in Figure 1. Starting with the lowest level, these are Observing Systems, Data Assembly Centers, Data Access Services, Utility Services, and User Applications. Other IOOS literature [e.g., 2] has grouped the three middle layers into a single "Data Management and Communications (DMAC) subsystem."
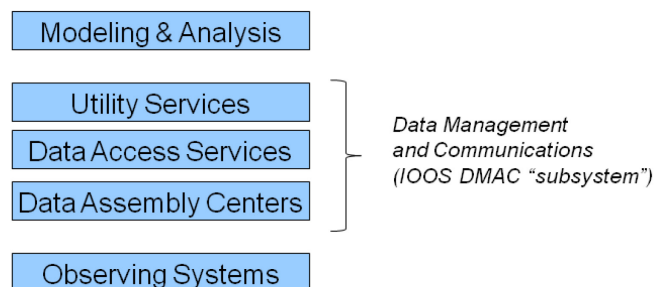


**Figure 1: IOOS architectural layers.**

Several cross-cutting themes must also be considered, as illustrated in Figure 2. These concerns apply mainly to DMAC, because IOOS does not have much control over the observing systems or the end-user applications.
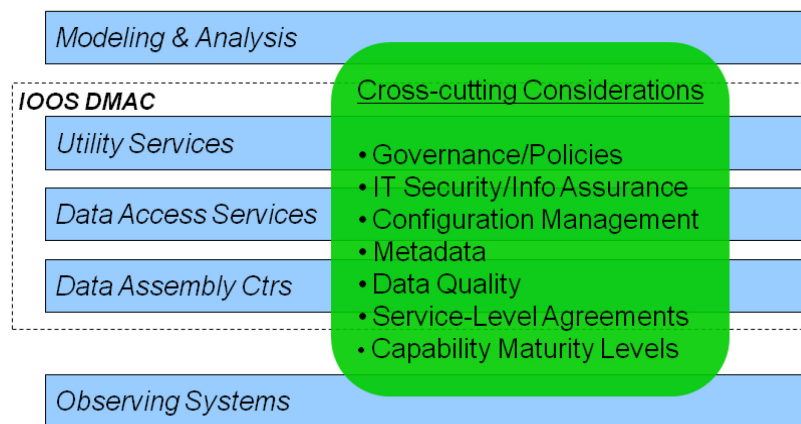
**Figure 2: Cross-cutting considerations for IOOS.**

- Governance includes the policies, standards, roles and responsibilities for stakeholders in this distributed, multi-agency effort.
- IT security includes at least the protection of individual servers in the system, and may also include methods to guard against data corruption.
- Configuration management is necessary to ensure that when functions are added or changed the various components that use that function can adapt or be kept in sync.
- Metadata must be collected at various points in the data life cycle by the appropriate personnel, and must then remain accessible and associated with the data.
- Data quality control may be performed at a variety of layers and components, but it is important to have an end-to-end capability for accessing data lineage and QA/QC metadata.
- Service-level agreements may be established with key data and service providers, specifying operational reliability, data quality and system performance requirements. However, it should also be permissible for non-operational service providers that support the same specifications to participate in the system.
- Capability maturity levels should be defined for various aspects of each service type, and made visible to the user of each service instance.

Figure 3 amplifies this five-layer model by populating some of the layers with components. This component view corresponds to the "Computational Viewpoint" from the ISO Reference Model for Open Distributed Processing (RM-ODP) [3]. The ellipsis (…) at the end of each layer in Figure 3 is meant to indicate that other components may belong on this diagram. The explanation below applies to the possible *future* scope of IOOS; in the next section we will use this figure to frame a description of progress to date.

- The Observing Systems layer represents both *in situ* and remote-sensing platforms, as well as other data collection methods such as trawl surveys, undersea imagery, and laboratory analysis of field samples.
- Data Assembly Centers (DACs) gather data from one or more observing platforms or measurement sources and then perform quality control, add metadata, store data at least temporarily, and send data to an archive for permanent storage. We also consider Archives and repositories of model outputs to be DACs. Such centers may be operated by federal, regional, international, commercial or academic entities.
- Data Access Services include standardized methods for requesting data. These services may be different for the various classes of data (e.g., individual observations vs. regularly-gridded data or imagery). Also, there may be different services for pulling data as needed, for subscribing to a feed of all data, or for receiving alerts based on predefined thresholds.
- Utility Services include functions such as Catalogs to help users find data; Registries of services, vocabularies and ontologies; conversion tools that translate between services or transform data from one representation into another; workflow services to orchestrate multi-step data processing or integration chains.
- Modeling and Analysis refers to any program or process that consumes ocean information. This layer includes numerical models for forecasting weather or ocean conditions, and scientific analysis software. It also includes simple web-based displays, commercial Geographic Information System (GIS) software, decision support tools, and system-monitoring software. "Client Applications" might be a better name, but "Modeling and Analysis" has been used extensively in IOOS documentation [e.g., 2].
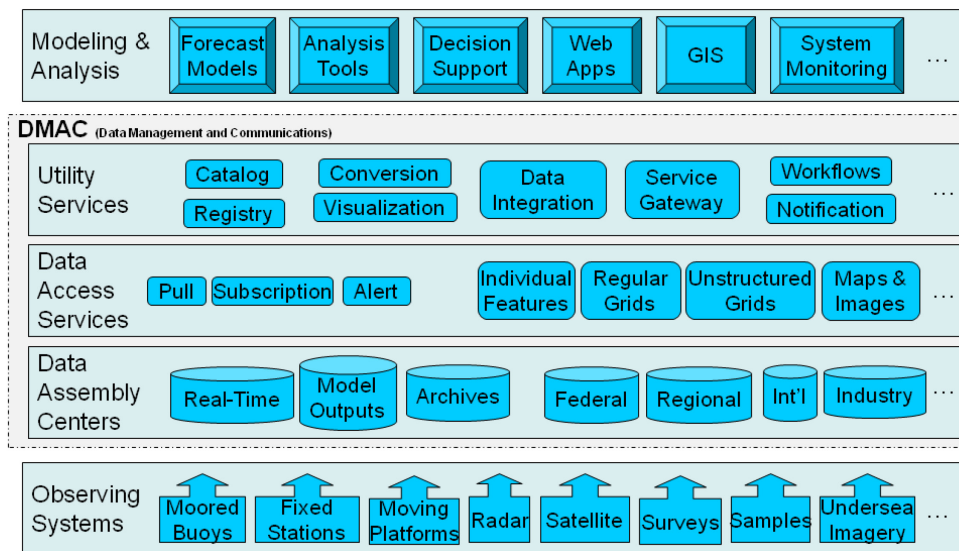
**Figure 3: Component types needed for IOOS.**

In some cases, data may flow from the observing system layer to the client via one or more of the other layers. More generally, however, it is important to note that the actual data flow may be much more complex: data branches to archives or models, which then become new sources of information; a workflow may have sub-components that ingest and transform data.

Many of the components shown in Figure 3 already exist. The challenge is to enable a more uniform view of a very heterogeneous set of components by establishing or leveraging a modest number of standardized services, formats and practices. IOOS takes a service-oriented architecture (SOA) approach, whereby existing systems are exposed via standard interfaces while, if desired, leaving legacy interfaces in place. IOOS also takes a federated approach, whereby data holdings remain with the stewards who are experts in those data types, rather than attempting to collect all ocean data into a single repository, and data providers can choose either to adopt standard practices or can arrange for another entity to do so on their behalf.

Figure 4 is a conceptual schematic of the federated, service-oriented architecture to be used by IOOS. Note that it is not IOOS-specific – the same basic architecture can be applied to many similar projects. The black horizontal lines represent the Internet, across which all communication between services occurs. Software can be programmed to request data and metadata from the standardized Data Access Services. Anything behind these access services can be left as-is as long as the service provides appropriate translation; data stewards maintain control over their own data and internal processes. In the upper left, we see that information sources (represented by diamond shapes) include observing systems or model outputs. A Data Assembly Center establishes one or more services to provide access to the information under its purview. If a data provider cannot establish a standard service, it can arrange for a Data Center to do so on its behalf (top center).

An IOOS Portal (Fig. 4, top right) will provide a web-based user interface to allow humans to search for and retrieve data without having to program specialized software. This Portal relies on a Registry that lists all the data services and a Catalog that routinely harvests the table of contents of each service to maintain an up-to-date list of available data. The Registry and Catalog can be accessed either through the Portal or by software applications that have been programmed to query them.

IOOS provides support for 11 regional associations (RAs) of coastal ocean observing systems. The bottom left of Figure 4 shows how an RA might establish its own portal to enable its own stakeholders to find and access regional models and observations. A Regional Portal may have a local Registry and Catalog, and it may provide standardized data access to the IOOS network on behalf of all the regional members it supports.

A Thematic Portal (bottom center) portal provides access to information corresponding to a particular theme or community of interest, rather than to a particular geographic region. Its Registry and Catalog will only list data and services of relevance to that theme.

Finally, the Utility Services (bottom left) discussed earlier may exist independent of any data center or portal, as shown in the bottom left, or may be offered on behalf of IOOS by an existing center.
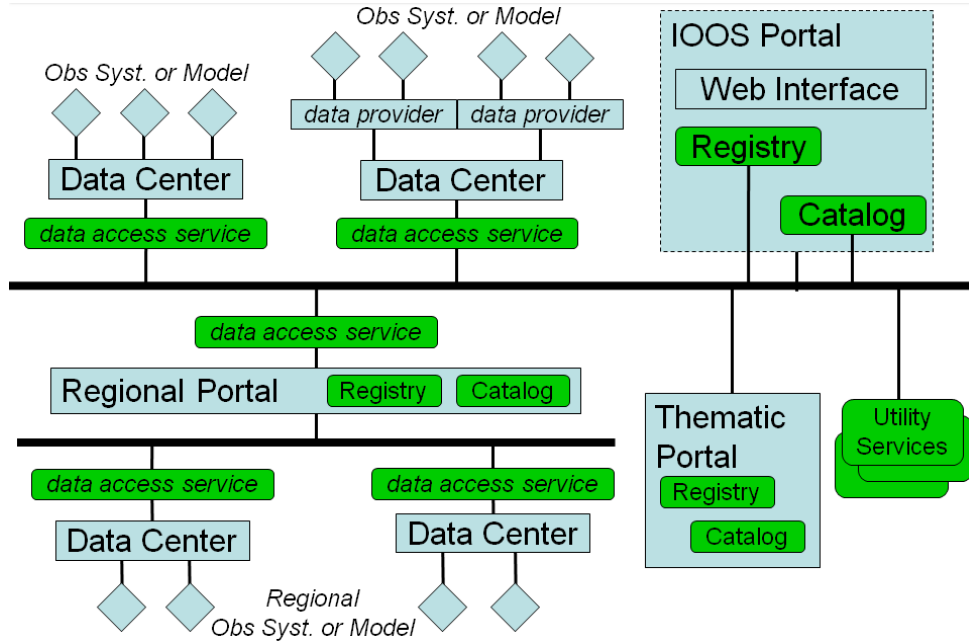
**Figure 2: IOOS federated, service-oriented architecture.**

### III. PROGRESS TO DATE

Progress towards IOOS by the NOAA IOOS Program Office has been mainly in three areas: (1) support for regional associations, including observing systems and getting at least a subset of their observations into national data assembly centers and world meteorological centers; (2) implementation of a Data Integration Framework (DIF) that addresses a subset of the architecture described above; and (3) planning, documentation and standardization activities in support of IOOS. We summarize the DIF project in this paper; further detail may be found in the proceedings of the previous MTS/IEEE Oceans conference [4].

The DIF is a limited-scope project to enable the evaluation of interoperability specifications, to demonstrate the feasibility and value of providing integrated ocean observations, and to provide the beginnings of initial operating capability for a nationwide IOOS data management capability. The initial scope of the DIF included three data assembly centers (DACs), four customers, and seven observed properties, but has since broadened as indicated below. In 2007, preparatory system engineering resulted in Functional Requirements [5] and Concept of Operations [6] documents. In 2008, establishment of this Data Integration Framework began in earnest with the implementation of at least one standardized data access service at each DAC.

The four official DIF customers are Harmful algal bloom forecasting, coastal inundation, hurricane intensification, and integrated ecosystem assessments. The customer projects are described in [7, this volume]. In the context of Figure 3, these customers are all Decision Support Tools or Forecast Models. Additional customer projects have now been initiated. In particular, we are formalizing arrangements with Google to provide data to the Ocean Observation layer in Google Earth, and for automatically providing event-specific packages of seismic data recorded by the DART buoys to tsunami scientists.

#### A. Data Access Services and Encoding Conventions

The Data Integration Framework project identified three general classes of scientific information to target first — *in situ* feature data, gridded coverage data, and images of data — and recommended a web service and encoding convention to be used in each case. These recommendations were intended to standardize a small number of data access methods and thereby to enable a single client application to obtain data from multiple providers, and to harmonize the representation of data from different providers. These services can be established either instead of or in addition to prior arrangements between individual providers and customers. The DIF services and encodings are summarized in Table 1 and described in more detail below. In the context of Figure 3, these data access services are all Pull services.

For *in situ* observations such as those from buoys, fixed sensors and volunteer observing ships, the DIF uses the Open Geospatial Consortium (OGC) Sensor Observation Service (SOS) [8] serving data and metadata encoded in Extensible Markup Language (XML). The XML employs a Geography Markup Language (GML) [9, 10] application schema based on the OGC Observations and Measurements (O&M) specification [11]. We implemented the SOS "core operations profile," which allows users to request data, observation metadata, or service metadata.

**Table 1: Web services and data encodings used in the IOOS Data Integration Framework.**

| Data Type | Web Service | Encoding |
|---|---|---|
| Feature collections (*in-situ* point or profile data, time series, trajectories) | Sensor Observation Service | Geography Markup Language |
| Regular grids (model outputs, Level 3 satellite data, radar surface currents) | Data Access Protocol or Web Coverage Service | Network Common Data Format |
| Georeferenced images of data | Web Map Service | Common image formats |

For serving gridded observations (including ocean color from satellites, surface currents from high-frequency radar, and model outputs), the DIF adopted both the Data Access Protocol (DAP) [12] and the OGC Web Coverage Service (WCS) [13]. Both protocols are suitable for accessing regular grids; DAP also supports irregular grids. WCS is explicitly called out in the GEOSS architecture and is supported by some commercial off-the-shelf (COTS) Geographic Information System (GIS) tools. DAP is well used in the NOAA scientific community and has been approved as recommended standard by the IOOS DMAC steering team. In practice, many data providers used a software package (THREDDS Data Server) that supports both DAP and WCS. The DIF recommends that gridded data be encoded in Network Common Data Form (NetCDF) [14] with Climate and Forecast (CF) conventions [15].

For images of data, the DIF recommends the OGC Web Map Service (WMS) [16], which generates georeferenced visualizations (i.e., "maps") upon request to the user's specifications. WMS is an OGC specification and an international standard (ISO 19128) [17].

*B.  Data Provider Implementations*

The primary DIF data assembly centers (DACs) are the National Data Buoy Center (NDBC) at NOAA's National Weather Service (NWS), the Center for Operational Oceanographic Products and Services (CO-OPS) in NOAA's National Ocean Service (NOS), and the CoastWatch program in NOAA's National Environmental Satellite Data and Information Service (NESDIS). In the context of Figure 3, NDBC, CO-OPS and CoastWatch are all considered real-time DACs: they provide access to current or recent observations.

The NOAA centers are all Federal (i.e., US government) DACs. We are beginning collaborations with other federal agencies including US Geological Survey, Environmental Protection Agency and US Army Corps of Engineers to standardize or interoperate with their data management practices.

NDBC assembles data from several buoy networks, including NWS meteorological platforms, the Deep-ocean Assessment and Report of Tsunamis (DART) warning buoys, the Tropical Atmosphere/Ocean (TAO) array for global climate studies, and a subset of observing platforms operated by the IOOS RAs. Real-time observations of 6 of the 7 core variables—ocean currents, temperature, salinity, water level, waves, and surface winds—from these buoys have been made accessible using Sensor Observation Service (SOS).

Also at NDBC, gridded surface currents computed from coastal high-frequency radar (HFR) observations of Doppler shifts have been made available using DAP/WCS. Images of the current vectors have been made available using WMS.

CO-OPS operates a variety of fixed stations as part of the National Water Level Observation Network (NWLON) and the Physical Oceanographic Real-Time System (PORTS). Real-time observations of 5 of the 7 core variables—ocean currents, temperature, salinity, water level, and surface winds—from these stations, as well as air temperature, barometric pressure, and water level predictions, have been made accessible via SOS.

At CoastWatch, gridded chlorophyll concentration derived from ocean color observations by the Moderate Resolution Imaging Spectroradiometer (MODIS) aboard the *Aqua* satellite have been made available using DAP/WCS.

Though not part of the original DIF scope, five of the 11 IOOS regional associations are establishing SOS to serve *in situ* observations from buoys or fixed stations.

In the realm of model outputs, the NOAA Coast Survey Development Laboratory has provided access to a limited subset of their modeled surface currents. The NOAA Atlantic Oceanographic and Meteorological Laboratory (AOML) has provided access to a limited subset of its model-derived synthetic temperature profiles. All of the IOOS RAs have established access to regional models. These model outputs are all gridded products served using DAP/WCS.

*C.  Metadata*

Metadata is one of the important cross-cutting considerations discussed earlier. Metadata of interest to IOOS includes descriptions of sensors, multi-sensor platforms, and networks thereof; non-sensor measurement procedures; quality control procedures and their results; and general information such as geographic coverage, temporal range, and thematic keywords. We

have not yet progressed far in this realm. NDBC and CO-OPS are in the process of enhancing the metadata they make available about *in situ* platforms and sensors. This information will be in SensorML [18] format and accessed via the SOS DescribeSensor operation. We are also defining minimum requirements for ISO 19115 metadata. Each data access service instance also has an associated service metadata record. We envision a linked set of metadata records at appropriate levels of detail that reference each other.

### D. *Utility Services*

The NOAA National Marine Fisheries Service (NMFS) Southwest Fisheries Service Center (SWFSC) Environmental Research Division (ERD) has developed (separately from IOOS) a data translation and visualization service known as ERDDAP. ERDDAP is able to access data in a variety of formats and protocols, and to transform those data on-the-fly to other formats or representations desired by the user. We have supported additional ERDDAP development, including enabling it to read and translate from the SOS implementations that use the DIF XML encoding specification. We have also funded a collaborative project with the National Science Foundation (NSF) Ocean Observatories Initiative Cyber Infrastructure (OOI/CI) effort to prototype a translation service hosted in the cloud (i.e., on commercial virtualized computing resources), in which ERDDAP was decomposed into independently scalable components. This project is described in [19, this volume].

IOOS has also supported a project to develop a service gateway between DAP and OGC web services. Neither of these utility services has yet been tied into a framework that would allow a user to invoke data from one service and select from available utility services to apply transformations before receiving the resulting product.

## IV. FUTURE WORK

Some of the cross-cutting considerations in Figure 2 have been addressed in part, and instances of some of the component types shown in Figure 3 have been implemented, but much remains to be done. In this section we summarize some of the future work we hope to undertake in the near term.

Evaluation: The DIF project plan calls for an evaluation of the results in 2010. We will assess the service types selected, the encoding conventions used, metadata responses, utility to customers, and lessons learned. This evaluation will guide some of the future work.

Observing systems: We have not yet addressed moving platforms, but hope to include at least Volunteer Observing Ship (VOS) data in the NDBC SOS in the coming year. All data considered so far have been sensor-based measurements of physical properties. We have not yet handled biological data, surveys, laboratory samples, or undersea imagery. We hope to begin addressing at least some biological data types in 2010. We plan to add additional physical variables from existing data providers. Also, we would like to add other sources of similar variables (e.g., satellite sea surface temperature and sea surface height to complement *in situ* temperature and water level).

Data assembly centers: Archives have yet to be handled explicitly. The federal observations, but not all of the regional observations and model outputs, are transmitted to the National Oceanographic Data Center (NODC) for permanent storage. Some NODC holdings can be retrieved via DAP; none yet via WCS or SOS. We would like to ensure that in future all IOOS-related observations and model outputs are archived, and to standardize interfaces for getting data into or out of the archives.

Coordination at the international level occurs within the framework of activities like GEOSS and OGC; no international data flows through IOOS DACs (except in the sense that DART and TAO buoys may be in international waters).

Data from some commercial oil and gas platforms in the Gulf of Mexico is served by NDBC, but there are no explicitly commercial IOOS data assembly centers at this time.

Data access services: We have not yet addressed mechanisms for subscriptions on an ad hoc, temporary or low-volume basis. Two existing data subscription services of importance are the World Meteorological Organization (WMO) Global Telecommunications System (GTS) and the Unidata Internet Data Distribution (IDD) network. GTS is used primarily as part of the mechanism to move data to and between data centers, and would lie within or below the Data Assembly Centers layer of Figure 3. IDD is an internet-based system used by some high-volume customers. We also have not yet implemented protocols such as the OGC Sensor Alert Service.

Utility Services: The limited scope of the DIF (3 providers, 4 customers) did not require a service registry or a catalog of data. However, we are now investigating registry and catalog options, including the Unidata THREDDS Catalog, leveraging

commercial search engine capabilities, and implementing an OGC Catalog Service. Also, we have registered the NOAA SOS services in the GEOSS service and component registry.

Models and Analysis: It is critical that we develop a basic portal capability to provide a visual catalog of our existing and future services. Also, we wish to collaborate with additional customers to solve their specific problems in an interoperable manner than can be reused by other customers.

## V. CONCLUSION

The NOAA IOOS program has made initiated the development of IOOS data management capabilities as part of the Data Integration Framework project to provide interoperable access to several sources of basic oceanographic data. The present scope of the effort is modest, focusing on a small number of variables and providers, but the methodology is generally applicable to a wide variety of observations, providers and customers. We will evaluate and expand on this effort in the coming years, and we welcome collaboration with other data providers, customer groups and interested stakeholders.

## ACKNOWLEDGMENTS

## REFERENCES

[1]  *Integrated Coastal and Ocean Observation System Act of 2009*, 33 U.S.C. § 3603(c)(2)(D), 2009.
[2]  S. Hankin *et al.*, *Data Management and Communications Plan for Research and Operational Integrated Ocean Observing Systems,* Ocean.US Publication No. 6, 2005.
[3]  International Organization for Standardization. *Information technology — Open Distributed Processing — Reference model: Overview,* ISO/IEC 10746-1: 1998.
[4]  J. de La Beaujardiere. "The NOAA IOOS Data Integration Framework: Initial Implementation Report," in *Proc. MTS/IEEE Oceans 2008 Conf.*, IEEE Marine Technical Society, paper #080515-116, 2008.
[5]  *Data Integration Framework (DIF) Functional Requirements Document,* NOAA Integrated Ocean Observing System Program Office, 2007.
[6]  *Data Integration Framework (DIF) Concept of Operations,* NOAA Integrated Ocean Observing System Program Office, 2008.
[7]  M. Weaks, S. Walker, and M. Koziara, "IOOS Customer-Focused Activities," in *Proc. MTS/IEEE Oceans 2009 Conf*, (this volume), IEEE Marine Technical Society, paper #090601-083, 2009.
[8]  A. Na and M. Priest, eds., *Sensor Observation Service,* version 1.0, Open Geospatial Consortium, 2007.
[9]  C. Portele, ed., *OpenGIS Geography Markup Language (GML) Encoding Standard,* version 3.2.1, Open Geospatial Consortium, 2007.
[10] *ISO 19136:2007, Geographic information — Geography Markup Language (GML),* International Organization for Standardization, 2007.
[11] S. Cox, ed. *Observations and Measurements – Part 1 - Observation schema,* version 1.0, Open Geospatial Consortium, 2007.
[12] P. Cornillon, J. Gallagher, and T. Skouros, "OPeNDAP: Accessing Data in a Distributed, Heterogeneous Environment," *Data Science Journal,* **2**, 5 Nov 2003, p. 164.
[13] A. Whiteside and J. Evans, eds., *Web Coverage Service (WCS) Implementation Standard,* version 1.1.2, Open Geospatial Consortium, 2008.
[14] R. Rew, G. Davis, S. Emmerson, H. Davies, and E. Hartne, *The NetCDF Users Guide,* version 4.0, Unidata Program Center, 2008.
[15] B. Eaton, J. Gregory, B. Drach, K. Taylor, S. Hankin, *NetCDF Climate and Forecast (CF) Metadata Conventions,* version 1.3, 2008.
[16] J. de La Beaujardière, ed., *OGC Web Map Service Interface,* version 1.3.0, Open Geospatial Consortium, 2004.
[17] *ISO 19128:2005, Geographic information — Web map server interface,* International Organization for Standardization, 2005.
[18] M. Botts and A. Robin, eds., *OpenGIS® Sensor Model Language (SensorML) Implementation Specification*, version 1.0, Open Geospatial Consortium, 2007.
[19] M. Arrott *et al.,* "Serving Ocean Model Data on the Cloud," in *Proc. MTS/IEEE Oceans 2009 Conf*, (this volume), IEEE Marine Technical Society, paper #090613-010, 2009.